

ORACLE INEQUALITY FOR A STATISTICAL RAUS–GFRERER TYPE RULE

QINIAN JIN* AND PETER MATHÉ†

Abstract. The authors study statistical linear inverse problems in Hilbert spaces. Approximate solutions are sought within a class of linear one-parameter regularization schemes, and the parameter choice is crucial to control the root mean squared error. Here a variant of the Raus–Gfrerer rule is analyzed, and it is shown that this parameter choice gives rise to error bounds in terms of oracle inequalities, which in turn provide order optimal error bounds (up to logarithmic factors). These bounds can only be established for solutions which obey a certain self-similarity structure. The proof of the main result relies on some auxiliary error analysis for linear inverse problems under general noise assumptions, and this may be interesting in its own.

Key words. statistical inverse problem, Raus–Gfrerer parameter choice, oracle inequality

AMS subject classifications. 47A52, secondary: 65F22, 65C60

1. Introduction. In this study we introduce a new parameter choice strategy for statistical linear inverse problems in Hilbert spaces. We consider the following linear equation

$$y^\delta = Tx^\dagger + \delta\xi, \quad (1.1)$$

where $T: X \rightarrow Y$ is a compact linear operator between Hilbert spaces X and Y , the parameter $\delta > 0$ denotes the noise level, and ξ stands for the additive noise, to be specified later as Gaussian white noise, which leads to observations y^δ . This is a standard model considered in statistical inverse problems. By using the singular system $\{s_j, u_j, v_j\}$ of T to write $Tx = \sum_j s_j \langle x, u_j \rangle v_j$, $x \in X$, the above model (1.1) is seen to be equivalent to the sequence space model

$$y_j^\delta = x_j + \delta\xi_j, \quad j = 1, 2, \dots,$$

with observations $y_j^\delta = \langle y^\delta, v_j \rangle / s_j$, the noise ξ_j is centered Gaussian with variance δ^2 / s_j^2 . The unknown solution x has coefficients x_j with respect to the basis u_j , $j = 1, 2, \dots$. This model is frequently analyzed, and we mention the recent survey [3]. In particular the minimax error is clearly understood if the solution sequence x_j , $j = 1, 2, \dots$ belongs to some Sobolev type ball. In particular, a series estimator $\hat{x}_k(y^\delta) = \sum_{j=1}^k c_j y_j^\delta$ (with appropriately chosen weights c_j) is (almost) optimal.

The important question is how to choose the truncation level (parameter, model) k based on the given data and the noise level δ . Parameter choice in statistical inverse problems, called *model selection* in this field, is an important issue, and we refer to [3] for a survey on this. Only recently, the *discrepancy principle*, which is the most prominent parameter choice in classical regularization theory, has been analyzed within the statistical context in [2]. Here, for any estimator $\hat{x} = \hat{x}(y^\delta)$ it requires to achieve that $\|T\hat{x} - y^\delta\| \asymp \delta$. Since the white noise ξ is not an element in Y , the discrepancy $\|T\hat{x} - y^\delta\|$ is not well-defined. Therefore, for statistical inverse problems, the traditional discrepancy principle can not be applied directly.

*Mathematical Sciences Institute, Bldg 27, Australian National University, ACT 0200 Australia (qinian.jin@anu.edu.au)

†Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, 10117 Berlin, Germany (peter.mathe@wias-berlin.de)

In order to make the discrepancy principle applicable to statistical inverse problems, we may consider, instead, the symmetrized equation with $A := T^*T \geq 0$ and $\zeta := T^*\xi$, as

$$z^\delta = T^*y^\delta = Ax^\dagger + \delta T^*\xi = Ax^\dagger + \delta\zeta. \quad (1.2)$$

Then, if the operator A has finite trace, the new misfit $\|A\hat{x} - z^\delta\|$ is almost surely finite, and it is tempting to require that

$$\|A\hat{x}(z^\delta) - z^\delta\| \asymp \delta, \quad (1.3)$$

which gives the discrepancy principle for the symmetrized equation. However, as was pointed out in [2], this plain use of the discrepancy principle leads only to suboptimal performance. Instead, the misfit $A\hat{x}(z^\delta) - z^\delta$ should be weighted, and if done accordingly, this can yield optimal rates of reconstruction. To be specific we consider the family of reconstructions

$$x_\alpha^\delta = (\alpha I + A)^{-1} T^*y^\delta, \quad \alpha > 0,$$

via Tikhonov regularization. The authors in [2] studied the *modified discrepancy principle*

$$\|(\lambda I + A)^{-1/2} (Ax_\alpha^\delta - z^\delta)\| \asymp \delta. \quad (1.4)$$

It is shown that an appropriate choice of $\lambda > 0$ yields order optimal reconstruction in many cases. However, the choice of λ requires the unknown smoothness of solution which makes the discrepancy principle into an *a priori* rule.

Instead, the authors in [12] considered the *varying discrepancy principle*

$$\|(\alpha I + A)^{-1/2} (Ax_\alpha^\delta - z^\delta)\| \asymp \delta \quad (1.5)$$

by relating $\lambda = \alpha$ in (1.4) to make the principle into an *a posteriori* one, and thus the weight depends on the parameter α under consideration. The main achievement in [12] is that this new principle may yield optimal order reconstruction (up to a logarithmic factor). However, it became transparent that such result holds only for solutions x^\dagger which satisfy certain self-similarity properties. This has an intuitive explanation: For large values of α , and this is where the discrepancy principle starts with, the misfit is dominated by the large singular numbers s_j . However, the approximation order is determined by all of the spectrum.

The varying discrepancy principle has another drawback. The regularization scheme, which is used to determine the candidate solutions x_α^δ must have higher qualification than given by the underlying smoothness in terms of general source conditions. For instance, if we use Tikhonov regularization, whose qualification is known to be 1, see [4], then the varying discrepancy principle gives order optimal reconstruction only for smoothness ‘up to 1/2’. This effect, which is inherent in the discrepancy principle in classical regularization context, is called *early saturation*, and it can be overcome by turning from the discrepancy principle to the so-called *Raus-Gfrerer rule* (RG-rule).

As Raus and Gfrerer proposed, instead of the discrepancy from (1.3) an additional weight should be used, which results in the RG-rule

$$\|(\alpha I + A)^{-1} (Ax_\alpha^\delta - z^\delta)\| \asymp \delta. \quad (1.6)$$

This is the starting point for the present study, the application of the RG-rule within the statistical context. It will be shown that an appropriate use of the RG-rule will yield order optimal results without the effect of early saturation. Actually, we will propose a statistical version of RG rule and establish some *oracle inequalities*, provided that the solution obeys some self-similarity. Oracle inequalities are widely used in statistics, see [3]. An oracle inequality guarantees that the estimator has a risk of the same order as that of the oracle. The oracle bound in particular implies that Tikhonov regularization can achieve order optimal reconstruction up to order 1.

This paper is organized as follows. We first precisely introduce the context, and then we state the main result with some discussion in Section 2. The proof of the main result will rely on preliminary results within the classical (deterministic noise) setting given in Section 3, however, under *general noise assumptions*. The results in this context may be interesting in their own. Finally, the proof of the main result is given in Section 4.

2. Setup and main result. We shall use the same setup as in [2, 12]. However, the parameter choice will be different.

2.1. Assumptions. We start with the description of the noise. We will mimic the notion of *Gaussian white noise* to the present case. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a (complete) probability space, and let \mathbb{E} be the expectation with respect to \mathbb{P} .

ASSUMPTION 2.1 (Gaussian white noise). *The noise $\xi = (\xi(y), y \in Y)$ in (1.1) is a stochastic process, defined on $(\Omega, \mathcal{F}, \mathbb{P})$ with the properties that*

1. *for each $y \in Y$ the random number $\xi(y) \in L_2(\Omega, \mathcal{F}, \mathbb{P})$ is a centered Gaussian random variable, and*
2. *for all $y, y' \in Y$ the covariance structure is $\mathbb{E}[\xi(y)\xi(y')] = \langle y, y' \rangle$.*

As a consequence, the mapping $y \rightarrow \xi(y)$ is linear, and we shall thus write $\xi(y) = \langle \xi, y \rangle$, we refer to [6] for details.

The related Gaussian process $\zeta := T^*\xi$ has covariance $\mathbb{E}[\langle \zeta, w \rangle \langle \zeta, w' \rangle] = \langle w, Aw' \rangle$, $w, w' \in X$ with the operator $A := T^*T$.

ASSUMPTION 2.2. *The operator A has finite trace $\text{Tr}[A] < \infty$.*

Under Assumption 2.2, Sazonov's Theorem, cf. [6], asserts that the element $\zeta := T^*\xi$ is a Gaussian random element in X (almost surely). Therefore the equation

$$z^\delta = Ax^\dagger + \delta\zeta \quad (2.1)$$

is a well defined linear equation in X (almost surely). This will be our main model from now on.

Moreover, Assumption 2.2 implies that the following function is well defined; for further properties we refer to [2].

DEFINITION 2.1 (effective dimension). *The function $\mathcal{N}(\lambda)$ defined as*

$$\mathcal{N}(\lambda) = \mathcal{N}_A(\lambda) := \text{Tr}[(A + \lambda I)^{-1}A], \quad \lambda > 0, \quad (2.2)$$

is called effective dimension of the operator A under white noise.

Along with the effective dimension, as in [12] we introduce the decreasing function $\varrho_{\mathcal{N}}(t)$ given by

$$\varrho_{\mathcal{N}}(t) := 1/\sqrt{t\mathcal{N}(t)}, \quad t > 0 \quad (2.3)$$

and its companion

$$\Theta_{\varrho_{\mathcal{N}}}(t) := t\varrho_{\mathcal{N}}(t), \quad t > 0. \quad (2.4)$$

The latter function is continuous and strictly increasing, hence its inverse is well-defined.

We recall the notion of linear regularization, see e.g. [5, Definition 2.2].

DEFINITION 2.2 (linear regularization). *A family of functions*

$$g_\alpha : (0, \|A\|] \mapsto \mathbb{R}, \quad 0 < \alpha \leq \|A\|,$$

is called regularization if they are piecewise continuous in α and the following properties hold:

1. *For each $0 < t \leq \|A\|$ we have that $|r_\alpha(t)| \rightarrow 0$ as $\alpha \rightarrow 0$;*
2. *There is a constant γ_1 such that $\sup_{0 \leq t \leq \|A\|} |r_\alpha(t)| \leq \gamma_1$ for all $0 < \alpha \leq \|A\|$;*
3. *There is a constant $\gamma_* \geq 1$ such that $\sup_{0 \leq t \leq \|A\|} \alpha |g_\alpha(t)| \leq \gamma_*$ for all $0 < \alpha < \infty$, where $r_\alpha(t) := 1 - tg_\alpha(t)$, $0 \leq t \leq \|A\|$, denotes the residual function.*

We further restrict the analysis to regularization schemes which are monotone

$$r_\alpha(t) \leq r_\beta(t), \quad \text{for } 0 < \alpha \leq \beta \quad (2.5)$$

and

$$0 \leq r_\alpha(t) \leq 1, \quad \text{for } \alpha > 0. \quad (2.6)$$

Hence Item (2) in Definition 2.2 holds with $\gamma_1 = 1$, and also $0 \leq tg_\alpha(t) \leq 1$. We also recall the following fact from [8, Lemma 2.3]: For $0 < \alpha \leq \beta$ there holds

$$0 \leq r_\beta(t) - r_\alpha(t) \leq (1 + \gamma_*) \frac{t}{\alpha + t} r_\beta(t). \quad (2.7)$$

Indeed, it follows from (2.5) and (2.6) that

$$0 \leq r_\beta(t) - r_\alpha(t) \leq (1 - r_\alpha(t))r_\beta(t) = tg_\alpha(t)r_\beta(t).$$

The result now follows from the observation that $(t + \alpha)g_\alpha(t) \leq 1 + \gamma_*$.

Having chosen an initial guess $x_0 \in X$ and a regularization g_α we construct the approximate solutions

$$x_\alpha^\delta := x_0 - g_\alpha(A)(Ax_0 - z^\delta), \quad \text{and} \quad x_\alpha := x_0 - g_\alpha(A)(Ax_0 - z);$$

for the noise free case we use $z := Ax^\dagger$. Recall that the element $\zeta = T^*\xi$ is a Gaussian random element in X (almost surely). Therefore, we will use the root mean squared error at a solution instance x^\dagger , given as

$$(\mathbb{E} [\|x^\dagger - x_\alpha^\delta\|^2])^{1/2}, \quad \alpha, \delta > 0. \quad (2.8)$$

2.2. Parameter choice. For the stopping criterion we will consider the following setup. Having chosen a constant $0 < q < 1$ we select the parameter α from the geometric family

$$\Delta_q := \{\alpha_k, \alpha_k := q^k \alpha_0, \quad k = 0, 1, 2, \dots\}. \quad (2.9)$$

For the statistical RG-rule we introduce the family of functions

$$s_\alpha(t) = \frac{\alpha}{t + \alpha}, \quad t, \alpha > 0, \quad (2.10)$$

which are the residual functions from Tikhonov regularization.

DEFINITION 2.3 (statistical RG-rule). *Given $\tau > 1$, $\eta > 0$ and $\kappa \geq 0$, let α_{RG} be the largest parameter $\alpha \in \Delta_q$ for which either*

$$\|s_\alpha(A)(Ax_\alpha^\delta - z^\delta)\| \leq \tau(1 + \kappa) \frac{\delta}{\varrho_{\mathcal{N}}(\alpha)}, \quad (2.11)$$

or

$$\Theta_{\varrho_{\mathcal{N}}}(\alpha) \leq \eta(1 + \kappa)\delta. \quad (2.12)$$

We will call the criteria (2.11) and (2.12) the regular stop and emergency stop, respectively. Notice that the regular stop in Definition 2.3 can be viewed as the Raus-Gfrerer rule applied to Lavrent'iev type regularization of the symmetrized equation (2.1).

2.3. Restricting the solution set. One important observation in the subsequent analysis, in particular in Section 3, will be that the RG-rule as introduced in §2.2 may fail for statistical problems (and also for bounded deterministic general noise), if the solution element x^\dagger has *abnormal* spectral behavior relative to the operator A . Therefore, we shall need the following restriction for the solution x^\dagger . To describe this we use the spectral resolution $(E_t)_{0 \leq t \leq \|A\|}$ of the (compact) non-negative self-adjoint operator A .

ASSUMPTION 2.3. *There exist $c_1 > 1$, $0 < c_2 < 1$ and $0 < t_0 < \|A\|$ such that*

$$\int_0^\alpha d\|E_t(x^\dagger - x_0)\|^2 \leq c_1^2 \int_{c_2\alpha}^\infty r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2$$

for all $0 < \alpha \leq t_0$.

The inequality in Assumption 2.3 with $c_2 = 1$ was introduced in [15] as a generalization of a restricted form on $x^\dagger - x_0$ in [10] for the (iterated) Tikhonov regularization.

EXAMPLE 2.1. For the n -times iterated Tikhonov regularization, we have $r_\alpha(t) = \alpha^n/(t + \alpha)^n$. It is easy to see that

$$|r_\alpha(t)| \geq c_3 \left(\frac{\alpha}{t}\right)^n \quad \text{for } t \geq c_2\alpha,$$

with $c_3 := (c_2/(1 + c_2))^n$. Therefore, in this case, Assumption 2.3 is equivalent to

$$\int_0^\alpha d\|E_t(x^\dagger - x_0)\|^2 \leq c_4 \alpha^{2n} \int_{c_2\alpha}^\infty t^{-2n} d\|E_t(x^\dagger - x_0)\|^2, \quad 0 < \alpha \leq t_0.$$

This, with $c_2 = 1$, is the condition used in [10].

EXAMPLE 2.2. For truncated singular value decomposition method we have

$$g_\alpha(t) = \begin{cases} 1/t, & t \geq \alpha, \\ 0, & t < \alpha \end{cases} \quad \text{and} \quad r_\alpha(t) = \begin{cases} 0, & t \geq \alpha, \\ 1, & t < \alpha. \end{cases}$$

Thus Assumption 2.3 becomes

$$\int_0^\alpha d\|E_t(x^\dagger - x_0)\|^2 \leq c_1 \int_{c_2\alpha}^\alpha d\|E_t(x^\dagger - x_0)\|^2, \quad \forall 0 < \alpha \leq t_0.$$

We observe that

$$\begin{aligned} \frac{\int_{c_2\alpha}^{\alpha} d\|E_t(x^\dagger - x_0)\|^2}{\int_0^{\alpha} d\|E_t(x^\dagger - x_0)\|^2} &= 1 - \frac{\int_0^{c_2\alpha} d\|E_t(x^\dagger - x_0)\|^2}{\int_0^{\alpha} d\|E_t(x^\dagger - x_0)\|^2} \\ &= 1 - \frac{\|E_{c_2\alpha}(x^\dagger - x_0)\|^2}{\|E_\alpha(x^\dagger - x_0)\|^2}. \end{aligned}$$

Therefore, for this scheme, Assumption 2.3 is equivalent to the existence of constants $0 < c_2 < 1$, $0 < \theta < 1$ and $0 < t_0 < \|A\|$ such that

$$\|E_{c_2\alpha}(x^\dagger - x_0)\| \leq \theta \|E_\alpha(x^\dagger - x_0)\|, \quad \forall 0 < \alpha \leq t_0. \quad (2.13)$$

EXAMPLE 2.3. For the asymptotical regularization we have $r_\alpha(t) = e^{-t/\alpha}$. Since $e^{-t/\alpha} \geq e^{-1}$ for $c_2\alpha \leq t \leq \alpha$, it is easy to see that Assumption 2.3 holds if (2.13) is satisfied.

EXAMPLE 2.4. For the Landweber iteration with $\|A\| = 1$, we have $r_\alpha(t) = (1 - t)^{[1/\alpha]}$, where $[1/\alpha]$ denotes the largest integer that is not greater than $1/\alpha$. Observing that for $0 < t \leq \alpha \leq 1/2$ there holds $(1 - t)^{[1/\alpha]} \geq (1 - t)^{1/\alpha} \geq (1 - \alpha)^{1/\alpha} \geq 1/4$. Therefore, Assumption 2.3 holds if (2.13) is satisfied.

2.4. Main result and discussion. The main result in this study is as follows.

THEOREM 2.1. *Let assumptions 2.1–2.3 hold. Let α_{RG} be chosen according to the statistical RG-rule with $\kappa = \sqrt{8|\log(1/\delta)|/\mathcal{N}(\alpha_0)}$. Then there is a constant C such that*

$$(\mathbb{E} [\|x^\dagger - x_{\alpha_{RG}}^\delta\|^2])^{1/2} \leq C \inf_{0 < \alpha \leq \alpha_0} \left\{ \|x_\alpha - x^\dagger\| + \frac{\delta(1 + \sqrt{|\log(1/\delta)|})}{\Theta_{\mathcal{N}}(\alpha)} \right\}.$$

The oracle inequality as established in Theorem 2.1 allows to state the error bound which is obtained under *known* general source condition and by an a priori parameter choice. We recall some notions.

DEFINITION 2.4 (general source set). *Given an index function ψ that is continuous, non-negative, and non-decreasing on $[0, \|A\|]$ with $\psi(0) = 0$, the set*

$$H_\psi := \{x \in X : x = \psi(A)v \text{ for some } \|v\| \leq 1\},$$

is called a general source set.

For solutions x^\dagger which belong to some source set, the bias $\|x_\alpha - x^\dagger\|$ can be bounded under the assumption that the chosen regularization has enough qualification, see e.g. [5]

DEFINITION 2.5 (qualification). *The regularization is said to have qualification ψ if there is a constant $\gamma < \infty$ such that*

$$|r_\alpha(t)| \psi(t) \leq \gamma \psi(\alpha), \quad \alpha > 0.$$

Notice that $x^\dagger - x_\alpha = r_\alpha(A)(x^\dagger - x_0)$. If the regularization has qualification ψ and $x^\dagger - x_0 \in H_\psi$, then

$$\|x^\dagger - x_\alpha\| \leq \gamma \psi(\alpha) \|v\| \leq \gamma \psi(\alpha).$$

By choosing $\alpha_\delta > 0$ to be the root of the equation

$$\Theta_{\varrho_N \psi}(\alpha) := \Theta_{\varrho_N}(t)\psi(t) = \delta \left(1 + \sqrt{|\log(1/\delta)|}\right),$$

we can use the the oracle inequality in Theorem 2.1 to obtain the following result.

COROLLARY 2.1. *Let the assumptions 2.1–2.3 hold, and let α_{RG} be chosen according to the statistical RG-rule with $\kappa = \sqrt{8|\log(1/\delta)|/\mathcal{N}(\alpha_0)}$. If the regularization has qualification ψ then*

$$\sup_{x^\dagger - x_0 \in H_\psi} \left(\mathbb{E} [\|x^\dagger - x_{\alpha_{RG}}^\delta\|^2] \right)^{1/2} \leq C\psi \left(\Theta_{\varrho_N \psi}^{-1} \left(\delta(1 + \sqrt{|\log(1/\delta)|}) \right) \right).$$

Thus, up to a logarithmic factor, the rate in Corollary 2.1 coincides with the one from [2, Theorem 1], which is known to be order optimal in many cases.

We conclude this section with an outline of the proof of Theorem 2.1. The basic idea is to reduce the argument to the one for bounded deterministic noise. The bound in Theorem 2.1 uses the effective dimension \mathcal{N} , or more precisely the function ϱ_N . This function naturally appears when considering the average performance of the noise under the weight $s_\alpha^{1/2}(A)$ because

$$\left(\mathbb{E} [\|s_\alpha^{1/2}(A)\zeta\|^2] \right)^{1/2} = (\text{Tr} [s_\alpha(A)A])^{1/2} = \sqrt{\alpha \mathcal{N}(\alpha)} = \frac{1}{\varrho_N(\alpha)}, \quad \alpha > 0. \quad (2.14)$$

Therefore, we choose a tuning parameter κ , as specified in Theorem 2.1, and define the set

$$Z_\kappa := \left\{ \zeta : \|s_\alpha^{1/2}(A)\zeta\| \leq (1 + \kappa) \frac{1}{\varrho_N(\alpha)}, \hat{\alpha} \leq \alpha \in \Delta_q \right\}, \quad (2.15)$$

where $\hat{\alpha}$ is the largest number in Δ_q satisfying

$$\Theta_{\varrho_N}(\hat{\alpha}) \leq \eta(1 + \kappa)\delta.$$

Let Z_κ^c denote the complement of Z_κ in X . Since $X = Z_\kappa \cup Z_\kappa^c$, we can use the Cauchy-Schwarz inequality to derive that

$$\left(\mathbb{E} [\|x^\dagger - x_\alpha^\delta\|^2] \right)^{1/2} \leq \sup_{\zeta \in Z_\kappa} \|x^\dagger - x_\alpha^\delta\| + \left(\mathbb{E} [\|x^\dagger - x_\alpha^\delta\|^4] \right)^{1/4} (\mathbb{P}[Z_\kappa^c])^{1/4}; \quad (2.16)$$

see [2, Proposition 3]. We will estimate the two terms on the right side of (2.16) with $\alpha = \alpha_{RG}$. Uniformly for $\zeta \in Z_\kappa$ the first term on the right can be considered as error estimate under bounded deterministic noise; and we will show in Section 3 that it can be bounded by the right hand side of the oracle inequality in Theorem 2.1. This analysis may be of independent interest. In Section 4 we will use some concentration inequality for Gaussian elements in Hilbert space to show that the second term on the right in (2.16) is negligible; this is enough for us to complete the proof of Theorem 2.1

3. Auxiliary results for bounded noise. The situation for bounded deterministic noise which resembles the Gaussian white noise case is regularization under some specifically chosen weighted noise. We recall the function s_α from (2.10). As could be seen from the set Z_κ in (2.15) the appropriate setup will be as follows.

ASSUMPTION 3.1. *There is a function $\alpha \rightarrow \delta(\alpha) > 0$ defined on $(0, \infty)$ that is non-decreasing, while $\alpha \rightarrow \delta(\alpha)/\sqrt{\alpha}$ is non-increasing such that the noise ζ obeys*

$$\delta \|s_\alpha^{1/2}(A)\zeta\| \leq \delta(\alpha), \quad \hat{\alpha} \leq \alpha \in \Delta_q, \quad (3.1)$$

where $\hat{\alpha} \in \Delta_q$ is the largest parameter such that $\hat{\alpha} \leq \eta\delta(\hat{\alpha})$ with $\eta > 0$ being a given small number.

Because $\alpha \rightarrow \delta(\alpha)/\sqrt{\alpha}$ is non-increasing and $\alpha \rightarrow \sqrt{\alpha}$ is strictly increasing, it is easy to see that $\hat{\alpha}$ is well-defined.

REMARK 3.1. The setup in Assumption 3.1 on noise covers a variety of cases which have been subsumed under the notion of *general noise assumptions*, we refer to [14, 1]. Specifically, let us consider the following situation. Suppose that the noise ζ allows for a noise bound for some parameter μ with

$$\|A^{-\mu}\zeta\| \leq 1. \quad (3.2)$$

In this case we can bound

$$\delta\|s_\alpha^{1/2}(A)\zeta\| \leq \delta\|s_\alpha^{1/2}(A)A^\mu\|\|A^{-\mu}\zeta\| \leq \|s_\alpha^{1/2}(A)A^\mu\|\delta.$$

It is easily verified that the operator norms $\|s_\alpha^{1/2}(A)A^\mu\|$ are uniformly bounded for $\alpha > 0$ if and only if $0 \leq \mu \leq 1/2$. In this range we easily obtain that

$$\|s_\alpha^{1/2}(A)A^\mu\| \leq \alpha^\mu, \quad \alpha > 0.$$

The two limiting cases are $\mu = 0$, where we assume $\|\zeta\| = \|T^*\xi\| \leq 1$ which corresponds to *large noise*, and $\mu = 1/2$, where we assume $\|A^{-1/2}\zeta\| = \|\xi\| \leq 1$ which corresponds to the usual noise assumption in linear inverse problems in Hilbert spaces. In any of the cases $0 \leq \mu \leq 1/2$ we get a bounding function $\delta(\alpha) = \delta\alpha^\mu$, which obeys the requirements made in Assumption 3.1.

Let $\hat{\alpha} \in \Delta_p$ be defined as in Assumption 3.1, i.e. $\hat{\alpha} \in \Delta_q$ is the largest parameter such that $\hat{\alpha} \leq \eta\delta(\hat{\alpha})$.

DEFINITION 3.1 (RG-rule). *Given $\tau > 1$ and $\eta > 0$, we define $\alpha_* \in \Delta_q$ to be the largest parameter such that*

$$\alpha_* \geq \hat{\alpha} \quad \text{and} \quad \|s_{\alpha_*}(A)(Ax_{\alpha_*}^\delta - z^\delta)\| \leq \tau\delta(\alpha_*); \quad (3.3)$$

if such α_ does not exist, we define $\alpha_* := \hat{\alpha}$.*

We notice that the norm in the above criterion can be rewritten as

$$\|s_\alpha(A)(Ax_\alpha^\delta - z^\delta)\| = \|s_\alpha(A)r_\alpha(A)(Ax_0 - z^\delta)\|.$$

3.1. Properties of the RG-rule. We give some technical consequences of the stopping criterion which will be used later.

LEMMA 3.1. *Let $\alpha \in \Delta_q$ be any parameter such that $\alpha > \alpha_*$. Then there holds*

$$\frac{\delta(\alpha)}{\alpha} \leq \frac{1}{\tau - 1} \|x_\alpha - x^\dagger\|.$$

Proof. Since $\alpha > \alpha_*$, by the definition of α_* we must have

$$\tau\delta(\alpha) \leq \|s_\alpha(A)r_\alpha(A)(Ax_0 - z^\delta)\|.$$

Therefore, it follows from Assumption 3.1 that

$$\begin{aligned} \tau\delta(\alpha) &\leq \|s_\alpha(A)r_\alpha(A)(z - z^\delta)\| + \|s_\alpha(A)r_\alpha(A)(Ax_0 - z)\| \\ &\leq \|s_\alpha^{1/2}(A)r_\alpha(A)\|\delta(\alpha) + \|s_\alpha(A)A\|\|x_\alpha - x^\dagger\|. \end{aligned}$$

Since $0 \leq s_\alpha^{1/2}(t)r_\alpha(t) \leq 1$ and $0 \leq s_\alpha(t)t \leq \alpha$, we have $\|s_\alpha^{1/2}(A)r_\alpha(A)\| \leq 1$ and $\|s_\alpha(A)A\| \leq \alpha$. Consequently

$$(\tau - 1)\delta(\alpha) \leq \alpha\|x_\alpha - x^\dagger\|,$$

which gives the estimate. \square

LEMMA 3.2. *Let the parameter α_* be chosen by the RG-rule in Definition 3.1. Then*

$$\|s_{\alpha_*}(A)r_{\alpha_*}(A)(Ax_0 - z)\| \leq \gamma_0\delta(\alpha_*),$$

where $\gamma_0 := \max\{1 + \tau, \eta\|x^\dagger - x_0\|\}$.

Proof. If $\alpha_* = \hat{\alpha}$, then it follows from the definition of $\hat{\alpha}$ that $\alpha_* \leq \eta\delta(\alpha_*)$. Consequently

$$\begin{aligned} \|s_{\alpha_*}(A)r_{\alpha_*}(A)(Ax_0 - z)\| &= \|s_{\alpha_*}(A)r_{\alpha_*}(A)A(x^\dagger - x_0)\| \\ &\leq \alpha_*\|x^\dagger - x_0\| \leq \eta\|x^\dagger - x_0\|\delta(\alpha_*). \end{aligned}$$

Otherwise we have that $\alpha_* > \hat{\alpha}$. Then by the definition of α_* we have

$$\begin{aligned} \|s_{\alpha_*}(A)r_{\alpha_*}(A)(Ax_0 - z)\| &\leq \|s_{\alpha_*}(A)r_{\alpha_*}(A)(z - z^\delta)\| + \|s_{\alpha_*}(A)r_{\alpha_*}(A)(Ax_0 - z^\delta)\| \\ &\leq \|s_{\alpha_*}^{1/2}(A)r_{\alpha_*}(A)\|\delta(\alpha_*) + \tau\delta(\alpha_*) \leq (1 + \tau)\delta(\alpha_*), \end{aligned}$$

and the proof is complete. \square

3.2. Auxiliary inequalities: The impact of Assumption 2.3. The following inequalities may be of general interest. The first one goes back to [7, 9], see also [8, Lemma 2.4].

LEMMA 3.3. *For $0 < \alpha \leq \beta$ we have*

$$\|x_\beta - x_\alpha\| \leq \frac{1 + \gamma_*}{\sqrt{\alpha}} \|A^{1/2}s_\beta^{1/2}(A)r_\beta(A)(x^\dagger - x_0)\|.$$

Proof. We first notice that $x_\beta - x_\alpha = (r_\beta(A) - r_\alpha(A))(x^\dagger - x_0)$. The bound established in (2.7) yields that

$$\begin{aligned} \|x_\beta - x_\alpha\| &= \|(r_\beta(A) - r_\alpha(A))(x^\dagger - x_0)\| \\ &\leq (1 + \gamma_*)\|A(\alpha + A)^{-1}r_\beta(A)(x^\dagger - x_0)\| \\ &= \frac{1 + \gamma_*}{\alpha} \|As_\alpha(A)r_\beta(A)(x^\dagger - x_0)\|. \end{aligned}$$

We may write

$$As_\alpha(A) = A^{1/2}s_\alpha^{1/2}(A)\frac{1}{s_\beta^{1/2}}(A)s_\alpha^{1/2}(A)A^{1/2}s_\beta^{1/2}(A).$$

Observing that $0 \leq s_\alpha(t)t^{1/2} \leq \sqrt{\alpha}$ and $s_\alpha(t) \leq s_\beta(t)$ for $t \geq 0$, we have that $\|s_\alpha^{1/2}(A)A^{1/2}\| \leq \sqrt{\alpha}$ and $\|\frac{1}{s_\beta^{1/2}}(A)s_\alpha^{1/2}(A)\| \leq 1$. Therefore

$$\|As_\alpha(A)r_\beta(A)(x^\dagger - x_0)\| \leq \sqrt{\alpha}\|A^{1/2}s_\beta^{1/2}(A)r_\beta(A)(x^\dagger - x_0)\|,$$

which allows to complete the proof. \square

The bound from Lemma 3.3 does not suffice, and we need the following strengthening, where Assumption 2.3 is crucial.

LEMMA 3.4. *Suppose that Assumption 2.3 holds true. Then there is a constant $C < \infty$ such that for $0 < \alpha \leq \alpha_0$ there holds*

$$\|A^{1/2}s_\alpha^{1/2}(A)r_\alpha(A)(x^\dagger - x_0)\| \leq \frac{C}{\sqrt{\alpha}}\|s_\alpha(A)r_\alpha(A)A(x^\dagger - x_0)\|.$$

Proof. We use spectral calculus to write

$$\|A^{1/2}s_\alpha^{1/2}(A)r_\alpha(A)(x^\dagger - x_0)\|^2 = I_1(\alpha) + I_2(\alpha),$$

where

$$\begin{aligned} I_1(\alpha) &:= \int_0^\alpha ts_\alpha(t)r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2 \\ I_2(\alpha) &:= \int_\alpha^\infty ts_\alpha(t)r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2. \end{aligned}$$

We first bound I_2 . For $t \geq \alpha$ we have that $\alpha(t + \alpha) \leq 2\alpha t$, thus $1 \leq \frac{2}{\alpha}ts_\alpha(t)$, yielding

$$I_2(\alpha) \leq \frac{2}{\alpha} \int_\alpha^\infty t^2 s_\alpha^2(t) r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2 \leq \frac{2}{\alpha} \|s_\alpha(A)r_\alpha(A)A(x^\dagger - x_0)\|^2.$$

To estimate $I_1(\alpha)$ we will use Assumption 2.3. We will consider two cases: $0 < \alpha \leq t_0$ and $t_0 < \alpha \leq \alpha_0$.

When $0 < \alpha \leq t_0$, we use Assumption 2.3 to obtain from $ts_\alpha(t) \leq \alpha$ that

$$I_1(\alpha) \leq \alpha \int_0^\alpha d\|E_t(x^\dagger - x_0)\|^2 \leq c_1^2 \alpha \int_{c_2\alpha}^\infty r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2.$$

Since $t/(t + \alpha) \geq c_2/(1 + c_2)$ for $t \geq c_2\alpha$, we further obtain

$$\begin{aligned} I_1(\alpha) &\leq \frac{c_1^2(1 + c_2)^2}{c_2^2\alpha} \int_{c_2\alpha}^\infty \frac{\alpha^2 t^2}{(t + \alpha)^2} r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2 \\ &= \frac{c_1^2(1 + c_2)^2}{c_2^2\alpha} \int_{c_2\alpha}^\infty s_\alpha^2(t) r_\alpha^2(t) t^2 d\|E_t(x^\dagger - x_0)\|^2 \\ &\leq \frac{c_1^2(1 + c_2)^2}{c_2^2\alpha} \|s_\alpha(A)r_\alpha(A)A(x^\dagger - x_0)\|^2. \end{aligned}$$

Now we consider the case $t_0 < \alpha \leq \alpha_0$. We write $I_1(\alpha) = I_1^{(1)}(\alpha) + I_1^{(2)}(\alpha)$, where

$$\begin{aligned} I_1^{(1)}(\alpha) &:= \int_0^{t_0} ts_\alpha(t)r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2, \\ I_1^{(2)}(\alpha) &:= \int_{t_0}^\alpha ts_\alpha(t)r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2. \end{aligned}$$

We can bound, by using Assumption 2.3, the term $I_1^{(1)}(\alpha)$ as

$$I_1^{(1)}(\alpha) \leq c_1^2 \alpha \int_{c_2 t_0}^\infty r_{t_0}^2(t) d\|E_t(x^\dagger - x_0)\|^2.$$

Since $t_0 \leq \alpha$ implies $r_{t_0}(t) \leq r_\alpha(t)$, we have

$$I_1^{(1)}(\alpha) \leq c_1^2 \alpha \int_{c_2 t_0}^{\infty} r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2.$$

Observing that for $t \geq c_2 t_0$ there holds $\frac{t}{t+\alpha} \geq \frac{t}{t+\alpha_0} \geq \frac{c_2 t_0}{c_2 t_0 + \alpha_0}$, we further obtain

$$\begin{aligned} I_1^{(1)}(\alpha) &\leq \left(\frac{c_2 t_0 + \alpha_0}{c_2 t_0} \right)^2 \frac{c_1^2}{\alpha} \int_{c_2 t_0}^{\infty} \frac{t^2 \alpha^2}{(t + \alpha)^2} r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2 \\ &= \frac{c_1^2}{\alpha} \left(\frac{c_2 t_0 + \alpha_0}{c_2 t_0} \right)^2 \int_{c_2 t_0}^{\infty} s_\alpha^2(t) r_\alpha^2(t) t^2 d\|E_t(x^\dagger - x_0)\|^2 \\ &\leq \frac{c_1^2}{\alpha} \left(\frac{c_1 t_0 + \alpha_0}{c_2 t_0} \right)^2 \|s_\alpha(A) r_\alpha(A) A(x^\dagger - x_0)\|^2. \end{aligned}$$

To bound $I_1^{(2)}$, we observe that for $t_0 \leq t \leq \alpha$ there holds $1 \leq \frac{\alpha_0 + t_0}{t_0 \alpha} t s_\alpha(t)$. Consequently

$$\begin{aligned} I_1^{(2)}(\alpha) &\leq \frac{\alpha_0 + t_0}{t_0 \alpha} \int_{t_0}^{\alpha} t^2 s_\alpha^2(t) r_\alpha^2(t) d\|E_t(x^\dagger - x_0)\|^2 \\ &\leq \frac{\alpha_0 + t_0}{t_0 \alpha} \|s_\alpha(A) r_\alpha(A) A(x^\dagger - x_0)\|^2. \end{aligned}$$

Combining the above estimates we therefore obtain the desired bound with $C = \left(2 + \frac{c_1^2(1+c_2)^2}{c_2^2} + \frac{\alpha_0 + t_0}{t_0} + c_1^2 \left(\frac{c_2 t_0 + \alpha_0}{c_2 t_0} \right)^2 \right)^{1/2}$. \square

We summarize the results from Lemma 3.3 and Lemma 3.4 as follows.

COROLLARY 3.1. *Let Assumption 2.3 hold. Then there is a constant $C < \infty$ such that for all $0 < \alpha \leq \beta \leq \alpha_0$ there holds*

$$\|x_\beta - x_\alpha\| \leq \frac{C}{\sqrt{\alpha\beta}} \|s_\beta(A) r_\beta(A) A(x^\dagger - x_0)\|.$$

3.3. Deterministic oracle inequality. In this section we state the main auxiliary result for bounded deterministic noise, as this seems to be of independent interest.

THEOREM 3.1. *Let the assumptions 2.3 and 3.1 hold, and let the parameter α_* be chosen by the RG-rule starting with α_0 . Then there holds the oracle inequality, i.e. there is a constant C such that*

$$\|x_{\alpha_*}^\delta - x^\dagger\| \leq C \inf_{0 < \alpha \leq \alpha_0} \left\{ \|x_\alpha - x^\dagger\| + \frac{\delta(\alpha)}{\alpha} \right\}. \quad (3.4)$$

Proof. We first derive some preparatory results. Observing that $x^\dagger - x_\alpha = r_\alpha(A)(x^\dagger - x_0)$, we have from (2.5) that

$$\|x^\dagger - x_\alpha\| \leq \|x^\dagger - x_\beta\|, \quad \forall 0 < \alpha \leq \beta. \quad (3.5)$$

By the conditions on g_α we have

$$\frac{g_\alpha(t)}{s_\alpha^{1/2}(t)} = \frac{1}{\sqrt{\alpha}} \sqrt{g_\alpha(t)} \sqrt{g_\alpha(t)(\alpha + t)} \leq \frac{\sqrt{\gamma_*(1 + \gamma_*)}}{\alpha}.$$

Therefore, with $c_* = \sqrt{\gamma_*(1 + \gamma_*)}$ we have

$$\begin{aligned} \|x^\dagger - x_\alpha^\delta\| &\leq \|x^\dagger - x_\alpha\| + \|g_\alpha(A) \left(\frac{1}{s_\alpha^{1/2}} \right) (A) \left[s_\alpha^{1/2}(A)(z - z^\delta) \right]\| \\ &\leq \|x^\dagger - x_\alpha\| + \frac{c_*}{\alpha} \|s_\alpha^{1/2}(A)\zeta\| \delta. \end{aligned} \quad (3.6)$$

It then follows from Assumption 3.1 that

$$\|x^\dagger - x_\alpha^\delta\| \leq \|x^\dagger - x_\alpha\| + c_* \frac{\delta(\alpha)}{\alpha}. \quad (3.7)$$

Next we will prove the oracle inequality in two steps. We first restrict the oracle bound to $\alpha \in \Delta_q$, and we show that

$$\|x_{\alpha_*}^\delta - x^\dagger\| \leq C \inf_{\alpha \in \Delta_q} \left\{ \|x_\alpha - x^\dagger\| + \frac{\delta(\alpha)}{\alpha} \right\}. \quad (3.8)$$

In this case we shall distinguish the cases $\alpha > \alpha_*$ and $\alpha \leq \alpha_*$, respectively.

Case $\alpha > \alpha_*$ We first have from (3.7), (3.5) and the monotonicity of $\alpha \rightarrow \delta(\alpha)$ that

$$\|x_{\alpha_*}^\delta - x^\dagger\| \leq \|x_{\alpha_*} - x^\dagger\| + c_* \frac{\delta(\alpha_*)}{\alpha_*} \leq \|x_\alpha - x^\dagger\| + c_* \frac{\delta(\alpha_*/q)}{\alpha_*}.$$

Since $\alpha, \alpha_* \in \Delta_q$, we have $\alpha_*/q \in \Delta_q$ and $\alpha \geq \alpha_*/q > \alpha_*$. Then we can conclude, by using Lemma 3.1 and (3.5), that

$$\|x_{\alpha_*}^\delta - x^\dagger\| \leq \|x_\alpha - x^\dagger\| + \frac{c_*}{q(\tau - 1)} \|x_{\alpha_*/q} - x^\dagger\| \leq \left(1 + \frac{c_*}{q(\tau - 1)} \right) \|x_\alpha - x^\dagger\|.$$

Case $\alpha \leq \alpha_*$ We actually use Assumption 2.3 and its consequences. Based on Corollary 3.1 and Lemmas 3.2 we conclude in this case that there is a constant $C < \infty$ with

$$\begin{aligned} \|x_{\alpha_*} - x^\dagger\| &\leq \|x_\alpha - x^\dagger\| + \frac{C}{\sqrt{\alpha\alpha_*}} \|s_{\alpha_*}(A)r_{\alpha_*}(A)(Ax_0 - z)\| \\ &\leq \|x_\alpha - x^\dagger\| + C\gamma_0 \frac{\delta(\alpha_*)}{\sqrt{\alpha\alpha_*}}. \end{aligned}$$

Consequently, we deduce, using the bound (3.7) and that $\alpha \rightarrow \delta(\alpha)/\sqrt{\alpha}$ is non-increasing, that

$$\begin{aligned} \|x_{\alpha_*}^\delta - x^\dagger\| &\leq \|x_{\alpha_*} - x^\dagger\| + c_* \frac{\delta(\alpha_*)}{\alpha_*} \leq \|x_\alpha - x^\dagger\| + C\gamma_0 \frac{\delta(\alpha_*)}{\sqrt{\alpha\alpha_*}} + c_* \frac{\delta(\alpha_*)}{\alpha_*} \\ &\leq (C\gamma_0 + c_*) \left(\|x_\alpha - x^\dagger\| + \frac{\delta(\alpha)}{\alpha} \right). \end{aligned}$$

Finally, we show the oracle inequality in its full generality. To this end, let $0 < \alpha \leq \alpha_0$ be any number. Then there is $j \in \mathbb{N}$ such that $\alpha_j < \alpha \leq \alpha_j/q$. By using (3.5), the fact that $\alpha \rightarrow \delta(\alpha)$ is increasing, and the fact that $\alpha \rightarrow \delta(\alpha)/\alpha$ is decreasing, we obtain

$$\begin{aligned} \|x_\alpha - x^\dagger\| + \frac{\delta(\alpha)}{\alpha} &\geq \|x_{\alpha_j} - x^\dagger\| + \frac{\delta(\alpha_j/q)}{\alpha_j/q} \geq q \left(\|x_{\alpha_j} - x^\dagger\| + \frac{\delta(\alpha_j)}{\alpha_j} \right) \\ &\geq q \inf_{\beta \in \Delta_q} \left\{ \|x_\beta - x^\dagger\| + \frac{\delta(\beta)}{\beta} \right\}. \end{aligned}$$

Since $0 < \alpha \leq \alpha_0$ is arbitrary, we obtain

$$\inf_{0 < \alpha \leq \alpha_0} \left\{ \|x_\alpha - x^\dagger\| + \frac{\delta(\alpha)}{\alpha} \right\} \geq q \inf_{\alpha \in \Delta_q} \left\{ \|x_\alpha - x^\dagger\| + \frac{\delta(\alpha)}{\alpha} \right\}.$$

The proof is therefore complete. \square

3.4. Discussion. (a) From Lemma 3.1 and (3.5) it follows that

$$\frac{\delta(\alpha_*/q)}{\alpha_*/q} \leq \frac{1}{\tau-1} \|x_{\alpha_*/q} - x^\dagger\| \leq \frac{1}{\tau-1} \|x_0 - x^\dagger\|.$$

Since $\alpha \rightarrow \delta(\alpha)$ is non-decreasing, we obtain

$$\frac{\delta(\alpha_*)}{\alpha_*} \leq \frac{q}{\tau-1} \|x_0 - x^\dagger\|.$$

If in the definition of $\hat{\alpha}$ we take $0 < \eta < \frac{\tau-1}{q\|x_0 - x^\dagger\|}$, then we always have $\alpha_* > \hat{\alpha}$. Therefore, the RG rule in Definition 3.1 simply reduces to the form: α_* is the largest parameter in Δ_q such that

$$\|s_{\alpha_*}(A)(Ax_{\alpha_*}^\delta - z^\delta)\| \leq \tau\delta(\alpha_*).$$

The oracle inequality in Theorem 3.1 still holds for this simplified parameter choice rule.

(b) The oracle inequality established in Theorem 3.1 can be used to yield error bounds when the solution x^\dagger has smoothness given in terms of *general source conditions*, i.e., if $x^\dagger - x_0$ belongs to some source set introduced in Definition 2.4. To see this, we assume that the regularization has qualification ψ as in Definition 2.5 and $x^\dagger - x_0 \in H_\psi$. We also assume, as introduced in Remark 3.1, that the noise can be bounded as $\|A^{-\mu}\zeta\| \leq 1$, which results in $\delta(\alpha) = \delta\alpha^\mu$ for $0 \leq \mu \leq 1/2$. Then, for the parameter α_* , determined by the RG rule in Definition 3.1, it follows from Theorem 3.1 that

$$\|x_{\alpha_*}^\delta - x^\dagger\| \leq C \inf_{0 < \alpha \leq \alpha_0} \left\{ \psi(\alpha) + \frac{\delta}{\alpha^{1-\mu}} \right\}.$$

Associated to the smoothness ψ , let $\Theta_{\mu,\psi}(t) := t^{1-\mu}\psi(t)$, $t > 0$, which is a strictly increasing function. Given $\delta > 0$ we assign $\alpha_\delta > 0$ such that $\Theta_{\mu,\psi}(\alpha_\delta) = \delta$. Then we can conclude that

$$\|x_{\alpha_*}^\delta - x^\dagger\| \leq C \left\{ \psi(\alpha_\delta) + \frac{\delta}{\alpha_\delta^{1-\mu}} \right\} \leq 2C\psi(\Theta_{\mu,\psi}^{-1}(\delta)),$$

which was shown to be order optimal for x^\dagger with the above smoothness in [14, Theorem 4]. Thus, the present results cover part of the analysis carried out in [14]; it extends the stopping criteria studied there to the RG-rule, and hence this relates to [13]. However, the above approach is limited. First, the case of *small noise*, i.e., when $-1/2 < \mu \leq 0$ cannot be covered. Secondly, the oracle inequality is seen to hold only for those solutions x^\dagger satisfying Assumption 2.3.

4. Proof of the main result. The proof of Theorem 2.1 will be carried out in several steps, similar to the one in the recent studies [2, 12]. Our starting point is the inequality (2.16). Recall that Z_κ is the set defined by (2.15), i.e. $Z_\kappa \subset X$ consists of those realizations of the noise ζ obeying Assumption 3.1 along the sequence $\alpha_0, \dots, \hat{\alpha}$ with

$$\delta(\alpha) := (1 + \kappa) \frac{\delta}{\varrho_{\mathcal{N}}(\alpha)}, \quad \alpha > 0, \quad (4.1)$$

where $\hat{\alpha}$ is the largest number in Δ_q satisfying

$$\Theta_{\varrho_{\mathcal{N}}}(\hat{\alpha}) \leq \eta(1 + \kappa)\delta, \quad (4.2)$$

According to the definition of α_{RG} we have $\alpha_{RG} \geq \hat{\alpha}$.

In order to estimate the first term on the right of (2.16) with $\alpha := \alpha_{RG}$, we observe that when $\zeta \in Z_\kappa$, the parameter α_* determined by the RG rule in Definition 3.1 with $\delta(\alpha)$ given by (4.1) is equal to the parameter α_{RG} determined by the statistical RG rule in Definition 2.3. Therefore we may use Theorem 3.1 to conclude

$$\sup_{\zeta \in Z_\kappa} \|x^\dagger - x_{\alpha_{RG}}^\delta\| \leq C \inf_{0 < \alpha \leq \alpha_0} \left\{ \|x^\dagger - x_\alpha\| + \frac{(1 + \kappa)\delta}{\Theta_{\varrho_{\mathcal{N}}}(\alpha)} \right\}. \quad (4.3)$$

In the following we will estimate the second term on the right side of (2.16) with $\alpha = \alpha_{RG}$. We need some auxiliary results.

LEMMA 4.1. *Let Assumptions 2.2 hold. Let $\hat{\alpha} = \alpha_0 q^{\hat{n}} \in \Delta_q$ be the largest parameter satisfying (4.2). Then there is a constant C such that*

$$\hat{n} \leq C(1 + |\log(1/\delta)|).$$

Proof. Since $\|A\| > 0$ is the first eigenvalue of A , it follows from the definition of $\mathcal{N}(\alpha)$ that

$$\mathcal{N}(\alpha) = \text{Tr}[(\alpha I + A)^{-1}A] \geq \frac{\|A\|}{\alpha + \|A\|} \geq \frac{\|A\|}{\alpha_0 + \|A\|}, \quad 0 < \alpha \leq \alpha_0.$$

Therefore, with $C_0 := \sqrt{(\alpha_0 + \|A\|)/\|A\|}$, we obtain

$$\Theta_{\varrho_{\mathcal{N}}}(\alpha) = \sqrt{\frac{\alpha}{\mathcal{N}(\alpha)}} \leq C_0 \alpha^{1/2}, \quad 0 < \alpha \leq \alpha_0.$$

According to the definition of $\hat{\alpha}$ we have

$$\Theta_{\varrho_{\mathcal{N}}}(\hat{\alpha}/q) > \eta(1 + \kappa)\delta \geq \eta\delta.$$

Consequently $C_0(\hat{\alpha}/q)^{1/2} \geq \eta\delta$ which implies the result. \square

We shall also use some prerequisites from Gaussian random elements in Banach spaces, and we recall the following results from [11, Lemma 3.1 & Corollary 3.2].

LEMMA 4.2. *Let Ξ be any Gaussian element in some Banach space. Then*

$$\mathbb{P}[\|\Xi\| > \mathbb{E}[\|\Xi\|] + b] \leq e^{-\frac{b^2}{2v^2}},$$

with $v^2 := \sup_{\|w\| \leq 1} \mathbb{E} [\langle \Xi, w \rangle]^2$. Moreover, for each $p > 1$ there is a constant C_p such that

$$\mathbb{E} [\|\Xi\|^p]^{1/p} \leq C_p \mathbb{E} [\|\Xi\|].$$

We apply Lemma 4.2 to $\Xi := s_\alpha^{1/2}(A)\zeta s_\alpha^{1/2}(A)T^*\xi$. For fixed $\alpha \in \Delta_q$ we denote

$$Z_{\kappa, \alpha} := \left\{ \zeta : \|s_\alpha^{1/2}(A)\zeta\| \leq (1 + \kappa) \frac{1}{\varrho_{\mathcal{N}}(\alpha)} \right\}.$$

COROLLARY 4.1. *For each $0 < \alpha \leq \alpha_0$ there holds*

$$\mathbb{P} [Z_{\kappa, \alpha}^c] \leq e^{-\frac{\kappa^2 \mathcal{N}(\alpha)}{2}} \quad \text{and} \quad \left(\mathbb{E} [\|s_\alpha^{1/2}(A)\zeta\|^4] \right)^{1/4} \leq C_4 \frac{1}{\varrho_{\mathcal{N}}(\alpha)}.$$

Proof. We first estimate $\mathbb{P}[Z_{\kappa, \alpha}^c]$. The expected norm of Ξ can be bounded, cf. (2.14), as

$$\mathbb{E} [\|s_\alpha^{1/2}(A)\zeta\|] \leq \left(\mathbb{E} [\|s_\alpha^{1/2}(A)\zeta\|^2] \right)^{1/2} = \frac{1}{\varrho_{\mathcal{N}}(\alpha)}. \quad (4.4)$$

For any $w \in X$ with $\|w\| \leq 1$, the weak second moments can be bounded from above by

$$\mathbb{E} [\langle \Xi, w \rangle]^2 = \mathbb{E} [\langle \xi, T s_\alpha^{1/2}(A)w \rangle]^2 = \|T s_\alpha^{1/2}(A)w\|^2 \leq \|T s_\alpha^{1/2}(A)\|^2 \leq \alpha.$$

Thus we may apply Lemma 4.2 with $b := \kappa/\varrho_{\mathcal{N}}(\alpha)$ to conclude that

$$\mathbb{P} [Z_{\kappa, \alpha}^c] \leq e^{-\frac{\kappa^2}{2\alpha\varrho_{\mathcal{N}}^2(\alpha)}} = e^{-\frac{\kappa^2 \mathcal{N}(\alpha)}{2}},$$

which completes the proof of the first assertion. The second one is a consequence of (4.4) and Lemma 4.2. \square

Finally we turn to the proof of the main result.

Proof of Theorem 2.1. We will use (2.16) with $\alpha = \alpha_{RG}$. The first term on the right has been estimated in (4.3). By using Lemma 4.1 and Corollary 4.1 we obtain from $Z_\kappa^c = \bigcup_{\hat{\alpha} \leq \alpha \in \Delta_q} Z_{\kappa, \alpha}^c$ that

$$\mathbb{P} [Z_\kappa^c] \leq (\hat{n} + 1) \sup_{\hat{\alpha} \leq \alpha \in \Delta_q} \mathbb{P} [Z_{\kappa, \alpha}^c] \leq C (1 + |\log(1/\delta)|) e^{-\frac{\kappa^2 \mathcal{N}(\alpha_0)}{2}}.$$

For $\kappa = \sqrt{8|\log(1/\delta)|/\mathcal{N}(\alpha_0)}$ this yields

$$\mathbb{P} [Z_\kappa^c] \leq C (1 + |\log(1/\delta)|) \delta^4 \leq C \left(1 + \sqrt{|\log(1/\delta)|} \right)^4 \delta^4. \quad (4.5)$$

It remains to establish a bound for $\mathbb{E} [\|x^\dagger - x_{\alpha_{RG}}^\delta\|^4]$. We emphasize that the random element $x_{\alpha_{RG}}^\delta$ is no longer Gaussian in general, since the parameter α_{RG} depends on the data ζ . Hence we cannot apply Lemma 4.2 directly. Therefore we will use the error bound (3.6) which is valid for every ζ . By using the facts that $\alpha_{RG} \geq \hat{\alpha}$ and that the function $\alpha \mapsto s_\alpha^{1/2}(t)/\alpha$ is decreasing for each $t \geq 0$, we obtain

$$\|x^\dagger - x_{\alpha_{RG}}^\delta\| \leq \|x^\dagger - x_{\alpha_{RG}}\| + c_* \delta \frac{\|s_{\alpha_{RG}}^{1/2}(A)\zeta\|}{\alpha_{RG}} \leq \|x^\dagger - x_0\| + c_* \delta \frac{\|s_{\hat{\alpha}}^{1/2}(A)\zeta\|}{\hat{\alpha}}.$$

Since $\hat{\alpha}$ is deterministic, the element $s_{\hat{\alpha}}^{1/2}(A)\zeta$ is Gaussian. Thus we may use the bound on the fourth moment of $\|s_{\hat{\alpha}}^{1/2}(A)\zeta\|$ given in Corollary 4.1 to obtain

$$(\mathbb{E} [\|x - x_{\alpha_{RG}}^{\delta}\|^4])^{1/4} \leq \|x^{\dagger} - x_0\| + c_* C_4 \frac{\delta}{\Theta_{\varrho_{\mathcal{N}}}(\hat{\alpha})}.$$

Since the function $\alpha \rightarrow \varrho_{\mathcal{N}}(\alpha)$ is decreasing, it is easy to obtain that $\Theta_{\varrho_{\mathcal{N}}}(\hat{\alpha}) \geq q\Theta_{\varrho_{\mathcal{N}}}(\hat{\alpha}/q)$. By the definition of $\hat{\alpha}$ we then obtain $\Theta_{\varrho_{\mathcal{N}}}(\hat{\alpha}) \geq q\eta(1 + \kappa)\delta \geq q\eta\delta$. Consequently

$$(\mathbb{E} [\|x - x_{\alpha_{RG}}^{\delta}\|^4])^{1/4} \leq \|x^{\dagger} - x_0\| + \frac{c_* C_4}{q\eta}. \quad (4.6)$$

Combining the estimates (4.3), (4.5) and (4.6) with (2.16) and we use that $0 < \alpha \leq \alpha_0$ yields $\Theta_{\varrho_{\mathcal{N}}}(\alpha_0)/\Theta_{\varrho_{\mathcal{N}}}(\alpha) \geq 1$. We can conclude that

$$(\mathbb{E} [\|x^{\dagger} - x_{\alpha_{RG}}^{\delta}\|^2])^{1/2} \leq C \inf_{0 < \alpha \leq \alpha_0} \left\{ \|x_{\alpha} - x^{\dagger}\| + \frac{\delta(1 + \kappa)}{\Theta_{\varrho_{\mathcal{N}}}(\alpha)} \right\}.$$

The proof is therefore complete. \square

REFERENCES

- [1] Gilles Blanchard and Peter Mathé, Conjugate gradient regularization under general smoothness and noise assumptions, *J. Inverse Ill-Posed Probl.*, 18(6), 701–726, 2010.
- [2] ———, Discrepancy principle for statistical inverse problems with application to conjugate gradient regularization, Technical report, University of Potsdam, 2011.
- [3] L. Cavalier, Nonparametric statistical inverse problems, *Inverse Problems*, 24(3), 034004, 19pp, 2008.
- [4] H. W. Engl, M. Hanke and A. Neubauer, Regularization of Inverse Problems. Kluwer, Dordrecht, 1996.
- [5] Bernd Hofmann and Peter Mathé, Analysis of profile functions for general linear regularization methods, *SIAM J. Numer. Anal.*, 45(3), 1122–1141 (electronic), 2007.
- [6] I. A. Ibragimov and Y. A. Rozanov, *Gaussian random processes*, volume 9 of *Applications of Mathematics*, Springer-Verlag, New York, 1978. Translated from the Russian by A. B. Aries.
- [7] Qianjin Jin, On the iteratively regularized Gauss-Newton method for solving nonlinear ill-posed problems, *Math. Comp.*, 69(232), 1603–1623, 2000.
- [8] ———, On a class of frozen regularized Gauss-Newton methods for nonlinear inverse problems, *Math. Comp.*, 79(272), 2191–2211, 2010.
- [9] ——— and Zong-yi Hou, On an a posteriori parameter choice strategy for Tikhonov regularization of nonlinear ill-posed problems, *Numer. Math.*, 83(1), 139–159, 1999.
- [10] Stefan Kindermann and Andreas Neubauer, On the convergence of the quasioptimality criterion for (iterated) Tikhonov regularization, *Inverse Probl. Imaging*, 2(2), 291–299, 2008.
- [11] Michel Ledoux and Michel Talagrand, *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*, Springer-Verlag, Berlin, 1991. Isoperimetry and processes.
- [12] Shuai Lu and Peter Mathé, Varying discrepancy principle as an adaptive parameter selection in statistical inverse problems, *submitted*, 2012.
- [13] Peter Mathé and Ulrich Tautenhahn, Enhancing linear regularization to treat large noise, *J. Inverse Ill-Posed Probl.*, 19(6), 859–879, 2011.
- [14] ———, Regularization under general noise assumptions, *Inverse Problems*, 27(3), 035016, 2011.
- [15] Andreas Neubauer, The convergence of a new heuristic parameter selection criterion for general regularization methods, *Inverse Problems*, 24(5), 055005 (10pp), 2008.